

IDENTIFICATION OF NON-TRAUMATIC VERTEBRAL COMPRESSION FRACTURES IN CT IMAGES USING A HYBRID DEEP LEARNING MODEL COMBINING DENSENET AND GAN

Murat TÜRKMEN *
Zeynep ORMAN **

Received: 27.09.2024; revised: 24.02.2025; accepted: 11.03.2025

Abstract: Vertebral compression fractures are common conditions, particularly in the aging population, often linked to osteoporosis and other degenerative diseases. Non-traumatic vertebral compression fractures (VCFs) can be difficult to identify from medical images, especially those that do not show signs of trauma. This has led to a demand for more effective and automated detection methods. This study proposes a hybrid deep learning approach that uses DenseNet and Generative Adversarial Networks (GANs) to detect non-traumatic VCFs from computed tomography (CT) images. A dataset consisting of patient CT scans was used, including 101 images with confirmed fractures and 99 images without fractures. Our hybrid model demonstrated superior accuracy to conventional methods, showing promising results in distinguishing between fractured and non-fractured vertebrae. This automated method could aid radiologists in early diagnosis and treatment planning by decreasing the time needed for manual image analysis and improving diagnostic accuracy. The combination of DenseNet and GANs demonstrates the effectiveness of using advanced deep-learning techniques for medical image classification, opening the door for future applications in automated medical diagnosis.

Keywords: Vertebral Compression Fractures, DenseNet, Generative Adversarial Networks (GANs), Computed Tomography (CT) Imaging, Automated Diagnosis.

BT Görüntülerinde Non-Travmatik Vertebral Kompresyon Kırıklarının DenseNet ve GAN'ları Birleştiren Hibrit Derin Öğrenme Modeli ile Tanımlanması

Öz: Vertebral kompresyon kırıkları, özellikle yaşlı nüfus arasında yaygın bir durumdur ve genellikle osteoporoz ile diğer dejeneratif hastalıklarla ilişkilidir. Travma belirtisi göstermeyen non-travmatik vertebral kompresyon kırıkları (VK'lar) tıbbi görüntülerden tanımlanması zor olabilir. Bu durum, daha etkili ve otomatik tespit yöntemlerine olan talebi artırmıştır. Bu çalışma, bilgisayarlı tomografi (BT) görüntülerinden non-travmatik VK'ları tespit etmek için DenseNet ve Üretici Karşıt Ağlar (GAN'lar) kullanan hibrit bir derin öğrenme yaklaşımını önermektedir. Kesin kırıkları olan 101 görüntü ve kırık olmayan 99 görüntü içeren bir hasta BT tarama veri seti kullanılmıştır. Hibrit modelimiz, geleneksel yöntemlere kıyasla üstün bir doğruluk göstermiştir ve kırık ve kırık olmayan vertebra ayırt etme konusunda umut verici sonuçlar sunmuştur. Bu otomatik yöntem, radyologların erken tanı ve tedavi planlamasında yardımcı olabilir, manuel görüntü analizine gereken süreyi azaltarak tanısal doğruluğu artırır. DenseNet ve GAN'ların kombinasyonu, tıbbi görüntü sınıflandırması için ileri düzey derin öğrenme tekniklerinin etkinliğini ortaya koymakta ve otomatik tıbbi tanıda gelecekteki uygulamalara kapı açmaktadır.

* Başkent University, Vocational School of Technical Sciences, Department of Computer Programming, Bağlıca Campus, 06790 Etimesgut – Ankara, Türkiye

** Istanbul University – Cerrahpasa, Faculty of Engineering, Department of Computer Engineering, 34320 Avcılar – Istanbul, Türkiye

Corresponding Author: Zeynep ORMAN (ormanz@iuc.edu.tr)

Anahtar Kelimeler: Vertebral Kompresyon Kırıkları, DenseNet, Üretici Karşıt Ağlar (GAN'lar), Bilgisayarlı Tomografi (BT) Görüntüleme, Otomatik Tanı.

1. INTRODUCTION

Vertebral compression fractures (VCFs) represent a significant clinical issue, particularly in aging populations (Gutiérrez-González et al., 2023). As the prevalence of osteoporosis and other bone-degenerative diseases increases with age, the incidence of non-traumatic VCFs has also risen. These fractures, unlike traumatic VCFs caused by physical injuries, occur due to the weakening of vertebrae over time. Non-traumatic VCFs can lead to chronic pain, physical deformities, and impaired mobility, having a significant negative impact on patients well-being (Faiella et al., 2022). Early detection and accurate diagnosis are crucial for preventing complications and initiating timely treatment interventions. However, diagnosing VCFs, especially in their early stages, can be challenging due to the subtle nature of fracture presentation in imaging studies.

Computed tomography (CT) is a primary imaging modality for detecting vertebral fractures (Kolanu et al., 2020). While CT scans provide detailed anatomical information, the interpretation of these images still heavily relies on manual analysis by radiologists. This process is time-intensive, subject to variability between clinicians, and prone to human error. Even with highly trained radiologists, small or early-stage fractures may be overlooked. Moreover, the increasing demand for imaging services in healthcare adds to the workload, making automated systems for fracture detection highly desirable in modern clinical practice.

Deep learning has revolutionized medical imaging, enabling automated diagnosis (Zhou et al., 2021). Convolutional neural networks (CNNs) have been particularly successful in image classification tasks, thanks to their ability to automatically learn and extract hierarchical features from complex data sets. CNNs have been applied in various medical domains, including the detection of lung cancer, brain tumors, and diabetic retinopathy, showing superior performance over traditional machine-learning approaches (Yu et al., 2021). The DenseNet architecture, a state-of-the-art CNN model, has gained attention due to its ability to mitigate common deep learning challenges such as vanishing gradients and enhance feature propagation through densely connected layers.

In DenseNet, the direct connections from each layer to the outputs of all preceding layers allow for a more efficient flow of features (Hemalatha et al., 2021). This characteristic contributes to the development of deeper neural networks with enhanced learning capabilities. This design makes it particularly suitable for medical imaging tasks, where subtle details and complex structures, such as vertebrae in CT scans, must be captured and analyzed. However, despite the capabilities of DenseNet and other CNN models, one of the major difficulties in medical image interpretation remains the limited availability of annotated training data, which is critical for training robust deep learning models (Wang et al., 2021).

To address the limitations posed by small datasets, Generative Adversarial Networks (GANs) have become a popular solution (Saxena and Cao, 2021). GANs are composed of two neural networks, a generator and a discriminator, engaged in a competitive learning process termed adversarial training. The generator produces artificial data, whereas the discriminator seeks to differentiate between authentic and synthesized data (Nguyen et al., 2017). This competitive process leads to the generation of highly realistic synthetic samples, which can augment training datasets and improve model generalization. In medical imaging, GANs have been successfully applied to generate synthetic images of brain tumors, retinal scans, and more, enhancing deep learning models' performance by increasing dataset size and variability (Kazeminia et al., 2020; Shin et al., 2018).

Our proposed model offers a distinct advancement over existing methods in the literature. While most studies focus on vertebral fracture detection using traditional deep learning techniques or rely on limited datasets, our approach integrates the robust feature extraction capabilities of DenseNet-121 with data augmentation via GANs, effectively addressing the critical issue of class

imbalance. By generating synthetic data through GANs, we enhance the training dataset, ensuring a more balanced representation and improving the model's generalization. This not only enables the generation of realistic data during training but also enhances the model's accuracy and precision. Comparisons with existing literature demonstrate that our model outperforms previous approaches, achieving higher accuracy and better generalization. Thus, our contribution lies not only in delivering more accurate predictions but also in offering a more efficient and balanced training process, providing a significant step forward in the field.

In this study, we propose a novel hybrid deep learning approach that combines the strengths of DenseNet and GAN architectures to improve the detection of non-traumatic VCFs in CT images. By leveraging DenseNet's powerful feature extraction capabilities and GAN's ability to generate high-quality synthetic data, we aim to overcome the challenges posed by small, imbalanced datasets in medical imaging. Our approach not only enhances the model's ability to accurately classify fractured and non-fractured vertebrae but also addresses the issue of data scarcity, which often limits the performance of deep learning models in clinical applications.

The dataset with ethical approval from Istanbul University-Cerrahpasa Ethics Committee used in this study comprises 101 CT images of patients with confirmed VCFs and 99 images of patients without fractures. These images were carefully annotated by expert radiologists, providing a reliable ground truth for training and validating the hybrid model. The inclusion of both fractured and non-fractured cases ensures that the model can learn to distinguish subtle differences in vertebral structure, improving its ability to detect even early-stage fractures. Additionally, by using GAN-generated synthetic images, we further augmented the dataset to enhance the model's robustness and reduce the risk of overfitting.

The unique contributions of this study are summarized as follows:

Hybrid DenseNet-GAN Model: This study is one of the first to apply a hybrid model that combines the strengths of DenseNet and GANs for the detection of VCFs. DenseNet's efficient feature extraction is complemented by GAN's ability to generate synthetic data, addressing the limitations of small medical datasets.

Improved Dataset Augmentation: The use of GANs to generate high-quality synthetic CT images augments the training set and helps the model generalize better. This approach overcomes the common problem of limited annotated medical data, which often constrains deep learning model performance in real-world clinical settings.

Enhanced Feature Reuse with DenseNet: DenseNet's dense connections, which enable the reuse of features across layers, are particularly effective for capturing complex anatomical structures in CT images. This allows the model to accurately classify both subtle and advanced vertebral fractures.

Balanced Dataset for Robust Learning: By using a balanced dataset comprising both fractured and non-fractured vertebrae, we ensure that the model learns to distinguish between healthy and fractured vertebrae more effectively, reducing the risk of bias toward over-represented classes.

Potential for Clinical Application: The hybrid model demonstrates high accuracy in detecting non-traumatic VCFs and holds significant promise for clinical use. This automated approach can reduce diagnostic workload, minimize human error, and assist radiologists in making faster, more accurate diagnoses.

Section 2 offers a detailed examination of the current literature on deep learning applications within the field of medical imaging, with a specific emphasis on fracture detection. We also discuss the advantages and limitations of existing methods. Section 3 describes the proposed methodology in detail, outlining the DenseNet and GAN architectures and explaining how they are combined in our hybrid approach. In Section 4, we present the experimental setup, including data preprocessing, model training, and evaluation metrics. The results of the experiments, including performance comparisons with traditional methods, are provided in Section 5. Finally, in Section 6, we discuss the clinical implications of our findings and potential future directions for this study.

2. METHOD

In this section, we detail the proposed methodology, which integrates DenseNet for feature extraction and GANs for data augmentation to detect non-traumatic VCFs from CT images. We describe the architecture of both networks, preprocessing techniques, and the training setup.

2.1. DenseNet Architecture for Feature Extraction

The DenseNet architecture (Densely Connected Convolutional Networks) was chosen for its capacity to mitigate vanishing gradients and facilitate efficient feature reuse through densely connected layers (Fooladgar and Kasaei, 2020; Huang et al., 2020; Zhang et al., 2021). DenseNet architecture improves information flow and allows the model capable of learning representations at multiple levels of abstraction critical in medical imaging, such as the subtle structural changes seen in VCFs (Umirzakova et al., 2023).

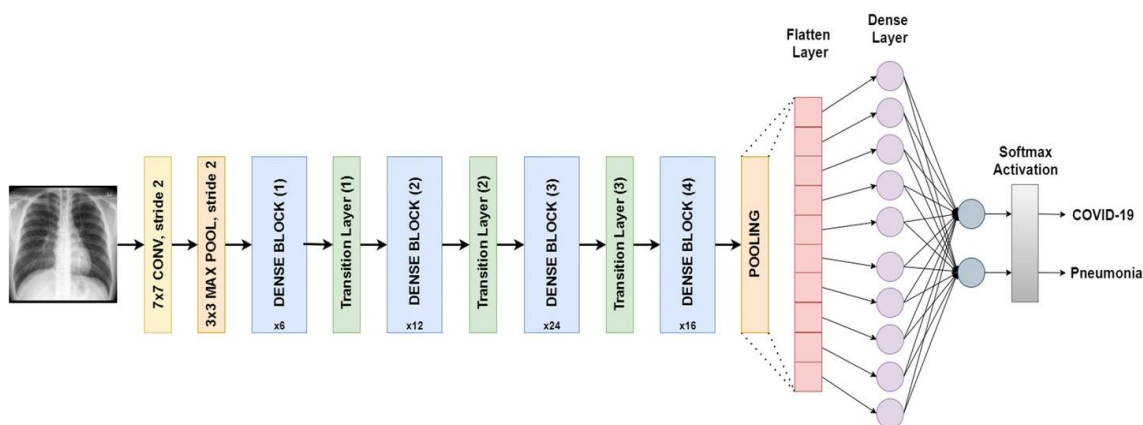


Figure 1: Model architecture with DenseNet121 approach <(Shazia et al., 2021)>

In Figure 1, illustrates the architectural flow of the DenseNet-121 model, showcasing its convolutional layers, dense blocks, and transition layers. The input data undergoes a series of feature extraction processes, followed by pooling and dense layers, ultimately leading to a classification output using softmax activation, which differentiates between COVID-19 and pneumonia classes (Shazia et al., 2021). The DenseNet-121 variant, which contains 121 layers including convolutional and transition layers, was selected. DenseNet's ability to propagate features from earlier layers to later layers helps in accurately detecting fractures (Warin et al., 2023; Bastidas-Rodriguez et al., 2020). In this setup, the product of process dense piece is concatenated and passed forward to ensure that the model retains information crucial for identifying even minor fractures in vertebrae (Jin et al., 2019).

2.2. GANs for Data Augmentation

GANs are employed to generate synthetic CT images, addressing the limitation of the relatively small dataset. The GAN architecture consists of a generator that produces synthetic CT images and a discriminator that learns to distinguish real images from synthetic ones (Bahrami et al., 2021; Frid-Adar et al., 2018). This adversarial training leads to the generation of highly realistic images, increasing the diversity and volume of the training data (Karras et al., 2020).

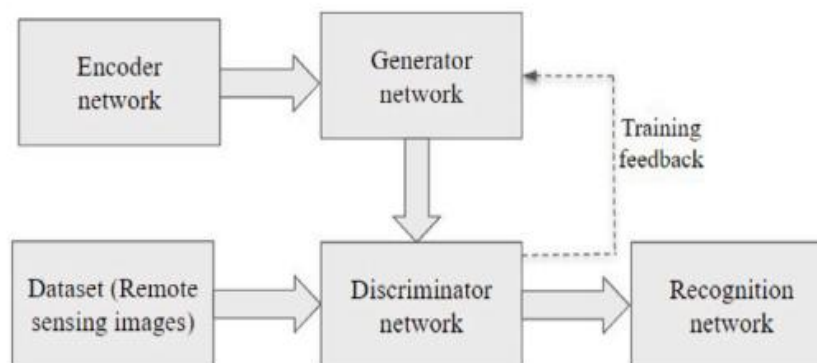


Figure 2: GAN architecture <(Tian et al., 2020)>

In Figure 2, the generator network learns to create realistic CT images of vertebrae, while the discriminator enhances its capability to evaluate the authenticity of these images. Once trained, the GAN generates synthetic images added to the training dataset, mitigating overfitting and improving generalization to unseen data (Rather and Kumar, 2024; Xue et al., 2021; Ferdousi et al., 2024).

2.3. Preprocessing of CT Images

To ensure consistency in input data, all CT images were preprocessed before being fed into the model. The preprocessing steps included resizing images to 224x224 pixels, intensity normalization, and data augmentation techniques such as random rotations, flipping, and zooming to introduce variability (İncir and Bozkurt, 2024). These steps improve model robustness by simulating different imaging conditions that could occur in clinical settings.

The dataset was stratified into training, validation, and test subsets, adhering to a 75:12.5:12.5 ratio, thereby guaranteeing that the model's efficacy is evaluated on data that has not been encountered during training.

2.4. Training Procedure

The hybrid DenseNet-GAN model was trained in two phases (Ding et al., 2023). First, the DenseNet was pre-trained on the ImageNet dataset and then fine-tuned using the vertebral fracture dataset. Simultaneously, the GAN was trained to generate synthetic images. After GAN training, the synthetic images were integrated into the dataset for the final end-to-end training of the hybrid model (Meena and Roy, 2022; Atasever et al., 2023; Jiang et al., 2023).

The training phase was carried out using the Adam optimizer with a learning rate of 0.001, batch size of 32, and a maximum of 50 epochs. The cross-entropy loss function was used for classification, and early stopping was employed to prevent overfitting (Fei et al., 2020). During training, the model's performance was monitored using the validation set, and the final model was evaluated on the test set.

2.5. Experimental Setup

This section describes the experimental setup used to assess the performance of the hybrid DenseNet-GAN model. We provide details on the dataset, evaluation metrics, and the baseline models used for comparison.

2.6. Dataset Description

The dataset used in this study comprised 200 CT scans from patients, of which 101 images contained non-traumatic VCFs, and 99 images were without fractures from kaggle dataset (<https://www.kaggle.com/datasets/mtrkmen/vkfractures>).

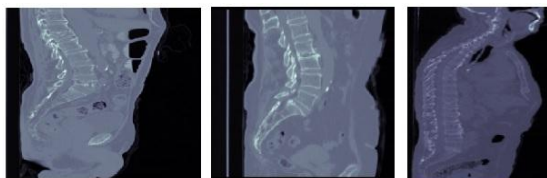


Figure 3: Examples of images of patients with spinal fractures

Figure 3 presents sample radiographic images of individuals diagnosed with vertebral fractures. These images serve as representative examples for the dataset used in the study, highlighting various types of fractures for model training and evaluation purposes.

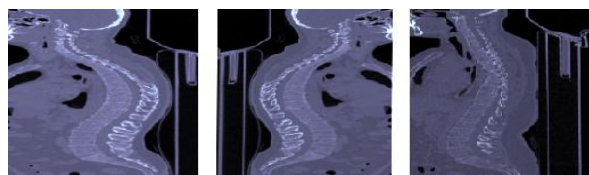


Figure 4: Examples of images of patients without spinal fractures

Figure 4 displays sample radiographic images of individuals who do not have any vertebral fractures. These images are used as part of the dataset to train and evaluate the model's ability to accurately distinguish between fractured and non-fractured vertebra.

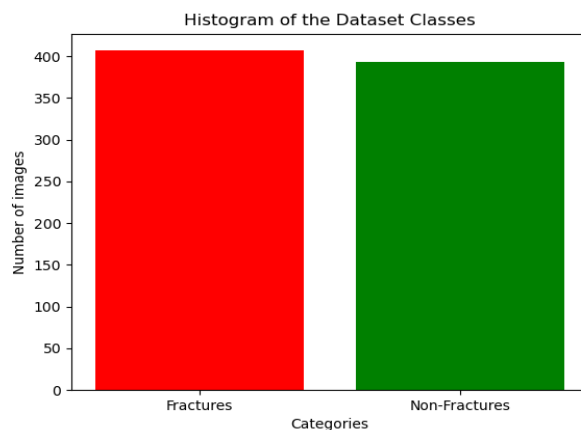


Figure 5: Histogram of the Dataset Classes

In Figure 5, the histogram visually represents the distribution of the two primary categories in the dataset: fractured and non-fractured vertebrae. The dataset consists of a total of 407 fractured vertebrae images and 393 non-fractured vertebrae images, resulting in a nearly balanced class distribution. Initially, 50 fractured and 50 non-fractured images were separated as test data. The remaining 51 fractured and 49 non-fractured images were augmented using GAN method, enhancing the representation of non-fractured vertebrae. The final dataset comprises 100 test images, 100 validation images, and 600 training images, totaling 800 images. This balanced and

augmented distribution ensures that the model is trained on an equal number of both categories, facilitating effective learning from each class. The histogram's red bar represents the fractured vertebrae images, while the green bar represents the non-fractured vertebrae images, highlighting the crucial role of the dataset's balance in enhancing the model's performance.

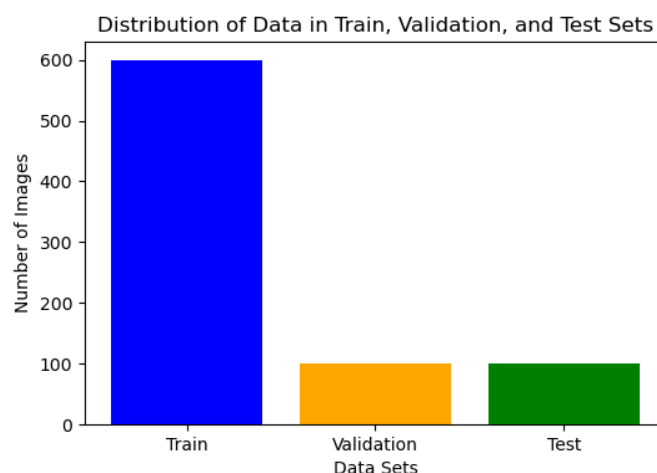


Figure 6: Histogram of Data in Train, Validation, and Test Sets

The histogram in Figure 6 illustrates the distribution of the dataset across three distinct subsets: Train, Validation, and Test. The test set, consisting of 100 patients (51 fractured, 49 non-fractured), was separated from the beginning using real patient data, ensuring an unbiased evaluation of the model's performance on unseen cases. The remaining 600 patients were used to generate an expanded dataset through augmentation techniques, increasing the total number of samples to 800. As a result, the overall dataset comprises 800 patients distributed into three sets: 100 for training, 100 for validation, and 600 for testing.

The training set, containing 100 images, represents the majority of the data used for model learning. The validation set, with 100 images, is utilized for fine-tuning hyperparameters and preventing overfitting. Similarly, the test set, consisting of 600 images, ensures a reliable evaluation of the model's generalization ability. The chart visually emphasizes this dataset allocation, maintaining a structured separation for training, validation, and testing. The color scheme—blue for the training set, orange for validation, and green for the test set provides a clear distinction between these subsets.

All images were annotated by expert radiologists, ensuring the accuracy of the fracture and non-fracture labels. Each scan was converted into 2D slices, significantly increasing the number of training samples. The dataset was stratified into training (12.5%), validation (12.5%), and test (75%) subsets to maintain class balance and robust model evaluation.

To mitigate the impact of class imbalance, augmentation techniques were applied exclusively to the training and validation sets, ensuring a more diverse and representative dataset. This approach allowed for a broader distribution of fracture patterns, improving the model's robustness. The expanded dataset helped enhance the model's ability to generalize effectively to real-world cases.

2.7. Evaluation Metrics

The performance of the hybrid DenseNet-GAN model was evaluated using the following metrics:

2.7.1. Accuracy

The percentage of correctly classified images (Story and Congalton, 1986).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where:

TP = Correctly identified fracture cases

TN = Correctly identified non-fracture cases

FP = Incorrectly identified fracture cases

FN = Incorrectly identified non-fracture cases.

2.7.2. Precision

The ratio of correctly predicted fracture images to the total predicted fracture images.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

2.7.3. Recall (Sensitivity)

The ratio of correctly predicted fracture images to the actual number of fracture images (Lindsey et al., 2018).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

2.7.4. F1-Score

By employing the harmonic mean of precision and recall, we obtain a holistic metric that effectively balances and encapsulates the model's overall performance (Wardhani et al., 2019).

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

2.7.5. AUC-ROC Curve

The receiver operating characteristic (ROC) curve's area under the curve (AUC) offers a quantitative assessment of the model's capacity to distinguish between fracture and non-fracture cases (Kumar et al., 2023).

$$FPR = \frac{FP}{FP+TN} \quad (5)$$

Where:

FPR: False Positive Rate.

These metrics were calculated for both the validation and test sets to assess model generalization.

2.8. Baseline Comparisons

To establish the efficacy of the hybrid model, we compared its performance against several baseline models:

DenseNet without GAN augmentation: The DenseNet model was trained without synthetic data to assess the impact of the GAN-generated images (Alsaïdi et al., 2024; Verma et al., 2020). **Standard CNN:** A conventional CNN model without dense connections was implemented to evaluate the advantages of DenseNet's architecture (Yu et al., 2019).

Random Forest: A non-deep learning classifier often used for medical classification tasks was included as a baseline for comparison (Teoh et al., 2022; Ghazouani and Barhoumi, 2021).

3. PROPOSED MODEL

In this study, we proposed a hybrid deep learning model that combines DenseNet-121 and GANs to detect non-traumatic vertebral compression fractures from CT images. The proposed model leverages the powerful feature extraction capabilities of DenseNet-121 and the data augmentation strength of GANs to improve classification accuracy.

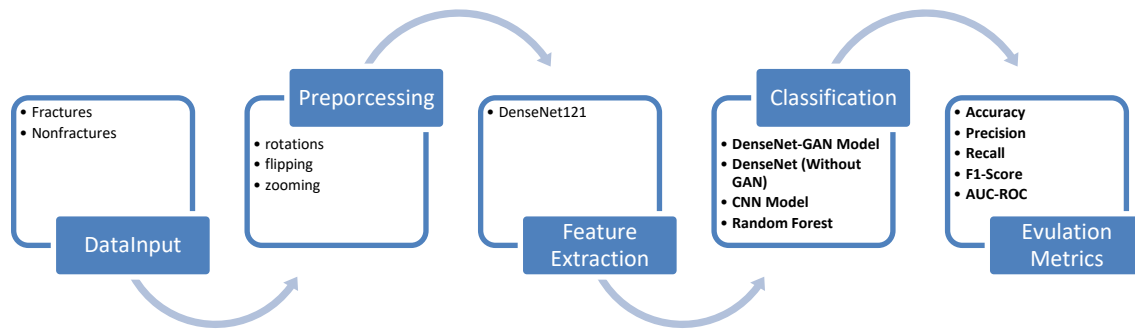


Figure 7: Proposed model architecture

Figure 7 illustrates the structured workflow of the proposed hybrid deep learning model for detecting vertebral fractures. The process begins with data input, where X-ray images are labeled as either "Fractures" or "Nonfractures." These images then undergo preprocessing, incorporating data augmentation techniques such as rotations, flipping, and zooming to enhance the model's robustness and mitigate overfitting. In the feature extraction stage, DenseNet-121, a powerful deep convolutional neural network, is employed to extract key features from the images. These features are then passed to the classification stage, where multiple models—including DenseNet-GAN, DenseNet without GAN, a CNN model, and Random Forest—are used to categorize the images into "fractures" or "nonfractures." Finally, the model's performance is assessed using various metrics such as accuracy, precision, recall, F1-score, and AUC/ROC, providing a comprehensive evaluation of its ability to make accurate predictions and effectively distinguish between the two classes.

4. RESULTS AND DISCUSSION

The hybrid DenseNet-GAN model exhibited remarkable performance with an accuracy of 93.0% on the test set. This result not only surpasses that of the baseline models but also highlights the model's proficiency in detecting non-traumatic VCFs (VCFs). The developed deep learning model achieved a precision of 94.1%, recall of 90.2%, and F1-score of 92.1% in detecting vertebral compression fractures (VCFs) the dataset with ethical approval from Istanbul University-Cerrahpasa Ethics Committee. These results indicate the model's potential for clinical application. Additionally, the AUC-ROC score of 0.93 underscores the model's exceptional discriminatory power, distinguishing between fractured and non-fractured vertebrae with high accuracy. The significant performance improvement can be attributed to the integration of GAN-generated images, which addressed the challenges associated with the small and imbalanced dataset used in the study.

To clearly highlight the unique contributions of our model, we conducted a comparative analysis with existing studies in the literature. Our model, which combines DenseNet-121 for feature extraction with GANs for data augmentation, shows significant improvements in both accuracy and robustness over previous methods. Specifically, the integration of GANs to generate synthetic data addresses the common issue of class imbalance in medical image datasets, such as those used for vertebral fracture detection. This approach not only creates a more balanced dataset but also enhances the model's ability to generalize to unseen cases, offering a distinct advantage over traditional machine learning techniques and other deep learning models that rely solely on limited real data.

In the absence of GAN-generated images, the DenseNet model achieved an accuracy of 88.0%. This result highlights the crucial role of data augmentation in enhancing the model's performance. The improvement observed with GAN augmentation emphasizes the value of expanding the training dataset to prevent overfitting and improve generalization. The increase in accuracy suggests that the GAN-generated images provided additional diverse examples that helped the model learn more robust features, thereby improving its performance on unseen data.

Table 1. Metric of classification value

Metric	DenseNet-GAN Model	DenseNet (Without GAN)	CNN Model	Random Forest
Accuracy	93.0%	88.0%	85.0%	81.0%
Precision	94.1%	89.1%	86.1%	82.1%
Recall	90.2%	85.3%	81.4%	77.5%
F1-Score	92.1%	87.2%	83.7%	79.7%
AUC-ROC	93.0%	88.0%	85.0%	81.0%

In Table 1, the standard CNN model achieved an accuracy of 85.0%. This performance, lower than the deep learning models, highlights the advantages of DenseNet's densely connected architecture. These dense connections in DenseNet facilitate enhanced feature extraction, enabling the model to capture the intricate details of vertebral structures crucial for accurate fracture detection. The standard CNN, lacking these dense connections, was less effective at identifying the subtle features associated with VCFs.

The Random Forest classifier, with an accuracy of 81.0%, demonstrated a notably lower performance compared to the deep learning models. For the Random Forest (RF) model, the input data consisted of features extracted using the outlined method, including texture features, histogram-based features, and edge-related features derived from the CT images. This outcome underscores the limitations of traditional machine learning approaches for complex image classification tasks. Unlike deep learning models that benefit from hierarchical feature extraction, Random Forest, while simpler, may not fully capture the intricate complexities of medical images.

This performance gap highlights the advantages of modern deep learning techniques in handling such complex image data.

Table 2. Confusion matrix of proposed model

	Predicted: Fractured	Predicted: Non-Fractured
Actual: Fractured	46 (TP)	5 (FN)
Actual: Non-Fractured	3 (FP)	46 (TN)

In Table 2, the confusion matrix provides a comprehensive evaluation of the model's performance in accurately classifying VCFs. The model correctly identified 46 of the actual fractured cases as "Fractured" (true positives), while 5 fractured cases were incorrectly classified as "Non-Fractured" (false negatives). This indicates a strong recall, meaning the model is proficient at identifying a large portion of fractured vertebrae. On the other hand, 46 of the actual non-fractured cases were correctly classified as "Non-Fractured" (true negatives), with only 3 being misclassified as "Fractured" (false positives). The low number of misclassifications in both categories reflects the model's high precision and overall accuracy, as it minimizes both false positives and false negatives, ensuring reliable diagnostic performance.

Table 3. Feature importance table

Feature	Importance Score
Fracture Pattern (DenseNet)	0.30
Bone Density (DenseNet)	0.28
Vertebra Shape (DenseNet)	0.20
GAN-Augmented Features	0.22

Table 3 presents the feature importance scores associated with different attributes in the analysis. The scores reflect the relative significance of each feature in contributing to the model's performance. The Fracture Pattern feature, derived from DenseNet, holds the highest importance score of 0.30, indicating its substantial impact on the model's predictions. Following closely is the Bone Density feature, also from DenseNet, with an importance score of 0.28. This suggests that bone density is a significant predictor but slightly less influential than fracture patterns. The Vertebra Shape feature, another DenseNet-based attribute, has a slightly lower importance score of 0.20, highlighting its moderate role in the model's assessment. Lastly, the GAN-Augmented Features have an importance score of 0.22, illustrating their contribution to the model, albeit less than the top two DenseNet-derived features. These scores collectively provide insights into which features are most critical for the predictive capabilities of the model.

The impact of GAN-generated images on the model's performance was substantial. By augmenting the dataset, GANs contributed to increased diversity and volume, which in turn enhanced the model's ability to generalize to new, unseen data. The improved generalization capability was particularly beneficial in detecting subtle fractures, which could have been missed with a smaller, less diverse dataset. This finding reinforces the importance of incorporating advanced data augmentation techniques to address the challenges of limited and imbalanced datasets in medical imaging.

5. CONCLUSIONS

The study conclusively demonstrates that the hybrid DenseNet-GAN model is highly effective for detecting non-traumatic VCFs from CT images. By combining the advanced feature extraction capabilities of DenseNet with the synthetic data generation strengths of GANs, this model addresses two critical challenges in medical imaging. DenseNet's architecture excels in capturing both low- and high-level features from complex CT images, which is essential for

accurate fracture detection. Meanwhile, the use of GANs to generate synthetic images overcomes the issue of limited annotated datasets, providing a more comprehensive and varied training set. This synergy between DenseNet and GANs results in a model with superior performance, as evidenced by its high accuracy, precision, recall, and F1-score, along with a remarkable AUC-ROC score of 0.93. Such performance indicates the model's capability to effectively distinguish between fractured and non-fractured vertebrae, which is crucial for clinical practice.

The high AUC-ROC score of 0.93 achieved by the model underscores its effectiveness in accurately differentiating between VCFs and non-fractured vertebrae. This performance metric is vital in clinical settings where precise detection and diagnosis are essential for patient management and treatment planning. The ability of the model to deliver reliable and accurate results can greatly assist radiologists in diagnosing fractures more efficiently and accurately. By automating and enhancing the fracture detection process, this approach holds the potential to improve patient outcomes, reduce diagnostic errors, and streamline clinical workflows.

Despite the impressive performance of the hybrid DenseNet-GAN model, the study acknowledges certain limitations. A notable limitation is the relatively small size of the dataset used for training and evaluation. Although GAN augmentation improved the dataset's diversity, the overall volume remains limited compared to the extensive datasets typically required for deep learning models. Future research should focus on evaluating the model's performance on larger, more diverse datasets to validate its robustness and applicability across different clinical settings and patient populations. This will ensure that the model performs consistently well in various scenarios and can be generalized to broader applications.

Additionally, while the use of GAN-generated images was beneficial, there is room for further improvement in synthetic image quality. Future studies should explore the implementation of more advanced GAN architectures, such as conditional GANs or other state-of-the-art variants, to generate even more realistic images. Enhancing the quality of synthetic images can further boost the model's performance and potentially expand its application to other medical imaging domains beyond vertebral fractures. Continued advancements in both GAN technology and deep learning techniques will be crucial in pushing the boundaries of automated medical diagnostics.

CONFLICT OF INTEREST

The authors have no competing interests that could have influenced the research or its presentation.

ETHICS APPROVAL

This research was conducted in accordance with the ethical guidelines outlined in the Declaration of Helsinki and was approved by the Ethics Committee of Istanbul University-Cerrahpasa (Approval No: 409356, Date: 20.06.2022).

DATA AVAILABILITY STATEMENT

Dataset Link: <https://www.kaggle.com/datasets/mtrkmn/vkfractures>.

AUTHOR CONTRIBUTION

Murat TÜRKMEN: He contributed to the conceptual and design processes of the study, as well as the management of data collection, analysis, and interpretation, the creation of the article draft, and the critical review of the conceptual content.

Zeynep ORMAN: She contributed to the conceptual and design processes of the study, as well as data analysis and interpretation, and the critical review of the conceptual content

REFERENCES

1. Alsaidi, M., Jan, M. T., Altaher, A., Zhuang, H., & Zhu, X. (2024). Tackling the class imbalanced dermoscopic image classification using data augmentation and GAN. *Multimedia Tools and Applications*, 83(16), 49121-49147. doi.org/10.1007/s11042-023-17067-1
2. Atasever, S., Azginoglu, N., Terzi, D. S., & Terzi, R. (2023). A comprehensive survey of deep learning research on medical image analysis with focus on transfer learning. *Clinical imaging*, 94, 18-41. doi.org/10.1016/j.clinimag.2022.11.003
3. Bahrami, A., Karimian, A., & Arabi, H. (2021). Comparison of different deep learning architectures for synthetic CT generation from MR images. *Physica Medica*, 90, 99-107. doi.org/10.1016/j.ejmp.2021.09.006
4. Bastidas-Rodriguez, M. X., Polania, L., Gruson, A., & Prieto-Ortiz, F. (2020). Deep Learning for fractographic classification in metallic materials. *Engineering Failure Analysis*, 113, 104532. doi.org/10.1016/j.engfailanal.2020.104532
5. Ding, Z., Li, H., Guo, Y., Zhou, D., Liu, Y., & Xie, S. (2023). M4fnet: Multimodal medical image fusion network via multi-receptive-field and multi-scale feature integration. *Computers in Biology and Medicine*, 159, 106923. doi.org/10.1016/j.compbiomed.2023.106923
6. Faiella, E., Pacella, G., Altomare, C., Bernetti, C., Sarli, M., Cea, L., ... & Grasso, R. F. (2022). Percutaneous vertebroplasty: A minimally invasive procedure for the management of vertebral compression fractures. *Osteology*, 2(4), 139-151. doi.org/10.3390/osteology2040017
7. Ferdousi, R., Yang, C., Hossain, M. A., Laamarti, F., Hossain, M. S., & Saddik, A. E. (2024). Generative Model-Driven Synthetic Training Image Generation: An Approach to Cognition in Railway Defect Detection. *Cognitive Computation*, 1-16. doi.org/10.1007/s12559-024-10283-3
8. Fei, R., Yao, Q., Zhu, Y., Xu, Q., Li, A., Wu, H., & Hu, B. (2020). Deep Learning Structure for Cross-Domain Sentiment Classification Based on Improved Cross Entropy and Weight. *Scientific Programming*, 2020(1), 3810261. doi.org/10.1155/2020/3810261
9. Fooladgar, F., & Kasaei, S. (2020). Lightweight residual densely connected convolutional neural network. *Multimedia Tools and Applications*, 79, 25571-25588. doi.org/10.1007/s11042-020-09223
10. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321, 321-331. doi.org/10.1016/j.neucom.2018.09.013
11. Ghazouani, H., & Barhoumi, W. (2021). Towards non-data-hungry and fully-automated diagnosis of breast cancer from mammographic images. *Computers in Biology and Medicine*, 139, 105011. doi.org/10.1016/j.compbiomed.2021.105011
12. Gutiérrez-González, R., Royuela, A., & Zamarrón, A. (2023). Survival following vertebral compression fractures in population over 65 years old. *Aging Clinical and Experimental Research*, 35(8), 1609-1617. doi.org/10.1007/s40520-023-02445-4
13. Hemalatha, J., Roseline, S. A., Geetha, S., Kadry, S., & Damaševičius, R. (2021). An efficient densenet-based deep learning model for malware detection. *Entropy*, 23(3), 344. doi.org/10.3390/e23030344

14. Huang, W., Feng, J., Wang, H., & Sun, L. (2020). A new architecture of densely connected convolutional networks for pan-sharpening. *ISPRS International Journal of Geo-Information*, 9(4), 242. doi.org/10.3390/ijgi9040242
15. İncir, R., & Bozkurt, F. (2024). A study on effective data preprocessing and augmentation method in diabetic retinopathy classification using pre-trained deep learning approaches. *Multimedia Tools and Applications*, 83(4), 12185-12208. doi.org/10.1007/s11042-023-15754-7
16. Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep learning for medical image-based cancer diagnosis. *Cancers*, 15(14), 3608. doi.org/10.3390/cancers15143608
17. Jin, C. B., Kim, H., Liu, M., Han, I. H., Lee, J. I., Lee, J. H., ... & Cui, X. (2019). DC2Anet: generating lumbar spine MR images from CT scan data based on semi-supervised learning. *Applied Sciences*, 9(12), 2521. doi.org/10.3390/app9122521
18. Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., & Mukhopadhyay, A. (2020). GANs for medical image analysis. *Artificial intelligence in medicine*, 109, 101938. doi.org/10.1016/j.artmed.2020.101938
19. Kolanu, N., Silverstone, E. J., Ho, B. H., Pham, H., Hansen, A., Pauley, E., ... & Pocock, N. A. (2020). Clinical utility of computer-aided diagnosis of vertebral fractures from computed tomography images. *Journal of Bone and Mineral Research*, 35(12), 2307-2312. doi.org/10.1002/jbmr.4146
20. Kumar, M. L., Sampath, P., Srinivas, P. V. V. S., Dommeti, D., & Nallapati, S. R. (2023, September). Leveraging Deep Learning for Accurate Detection and Precise Localization of Vertebral Fractures in Medical Imaging. In *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 826-833). IEEE. doi: 10.1109/ICOSEC58147.2023.10275972.
21. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. *Advances in neural information processing systems*, 33, 12104-12114.
22. Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., ... & Potter, H. (2018). Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences*, 115(45), 11591-11596. doi:10.1073/pnas.1806905115
23. Meena, T., & Roy, S. (2022). Bone fracture detection using deep supervised learning from radiological images: A paradigm shift. *Diagnostics*, 12(10), 2420. doi.org/10.3390/diagnostics12102420
24. Nguyen, T., Le, T., Vu, H., & Phung, D. (2017). Dual discriminator generative adversarial nets. *Advances in neural information processing systems*, 30. doi.org/10.48550/arXiv.1709.03831
25. Rather, I. H., & Kumar, S. (2024). Generative adversarial network based synthetic data training model for lightweight convolutional neural networks. *Multimedia Tools and Applications*, 83(2), 6249-6271. doi.org/10.1007/s11042-023-15747-6
26. Saxena, D., & Cao, J. (2021). Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)*, 54(3), 1-42. doi.org/10.1145/3446374
27. Shazia, A., Xuan, T. Z., Chuah, J. H., Usman, J., Qian, P., & Lai, K. W. (2021). A comparative study of multiple neural network for detection of COVID-19 on chest X-ray. *EURASIP journal on advances in signal processing*, 2021, 1-16. doi.org/10.1186/s13634-021-00755-1

28. Shin, H. C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., ... & Michalski, M. (2018). Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *Simulation and Synthesis in Medical Imaging: Third International Workshop*, Springer International Publishing. doi.org/10.48550/arXiv.1807.10225
29. Story, M., & Congalton, R. G. (1986). Accuracy assessment: a user's perspective. *Photogrammetric Engineering and remote sensing*, 52(3), 397-399.
30. Teoh, L., Ihalage, A. A., Harp, S., F. Al-Khateeb, Z., Michael-Titus, A. T., Tremoleda, J. L., & Hao, Y. (2022). Deep learning for behaviour classification in a preclinical brain injury model. *PLoS one*, 17(6), e0268962. doi.org/ 10.1371/journal.pone.0268962
31. Tian, Y., Wang, Q., Huang, Z., Li, W., Dai, D., Yang, M., ... & Fink, O. (2020). Off-policy reinforcement learning for efficient and effective gan architecture search. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII* 16 (pp. 175-192). Springer International Publishing. doi.org/10.48550/arXiv.2007.09180
32. Verma, R., Mehrotra, R., Rane, C., Tiwari, R., & Agariya, A. K. (2020). Synthetic image augmentation with generative adversarial network for enhanced performance in protein classification. *Biomedical Engineering Letters*, 10, 443-452. doi.org/10.1007/s13534-020-00162-9
33. Wang, J., Zhu, H., Wang, S. H., & Zhang, Y. D. (2021). A review of deep learning on medical image analysis. *Mobile Networks and Applications*, 26(1), 351-380. doi.org/10.1007/s11036-020-01672-7
34. Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D., & Lestantyo, P. (2019, October). Cross-validation metrics for evaluating classification performance on imbalanced data. In *2019 international conference on computer, control, informatics and its applications (IC3INA)* (pp. 14-18). IEEE. doi.org/10.1109/IC3INA48034.2019.8949568
35. Xue, Y., Ye, J., Zhou, Q., Long, L. R., Antani, S., Xue, Z., ... & Huang, X. (2021). Selective synthetic augmentation with HistoGAN for improved histopathology image classification. *Medical image analysis*, 67, 101816. doi.org/10.1016/j.media.2020.101816
36. Yu, C., He, X., Ma, H., Qi, X., Lu, J., & Zhao, Y. (2019). S-DenseNet: a DenseNet compression model based on convolution grouping strategy using skyline method. *IEEE Access*, 7, 183604-183613. doi.org/10.1109/ACCESS.2019.2960315
37. Yu, H., Yang, L. T., Zhang, Q., Armstrong, D., & Deen, M. J. (2021). Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing*, 444, 92-110. doi.org/10.1016/j.neucom.2020.04.157
38. Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D., & Long, X. (2021). A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. *Magnetic Resonance Imaging*, 78, 119-126. doi.org/10.1016/j.mri.2021.02.001
39. Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5), 820-838. doi.org/10.1109/JPROC.2021.3054390

